Assignment-1

Chirag, Charitha and Vijay

2023-01-28

Question 3

(a) Suppose the number of orders in Hall 4 Canteen in a particular weekday follows a Poisson distribution with mean 75. That in a weekend follows a Poisson distribution with mean 120. Setting seed = 1 and assuming the Hall 4 canteen was open in all 365 days in 2022, generate a (simulated) data set of daily orders in Hall 4 canteen. Treat this data as the population data, and calculate μ.

```
set.seed(1)
## function takes input as number of days or size
fx_poisson <- function(n)
{
  output <- array(0,dim = length(n))
  for(i in 1:n)
  {
    ##1st Jan 2022 was Saturday

    if(i %% 7 == 1 | i %% 7 == 2) #for weekends
    {
      output[i] <- rpois(1,120)
    }
    else  #for weekdays
    {
      output[i] <- rpois(1,75)
    }
  }
  return(output)
}
```

Calling function for 365 days of year 2022 and 'μ' equals to

```
## [1] 87.90411
```

(b) Take a sample with and without replacement of size n from this data set and find the mean $\bar{Y}$.

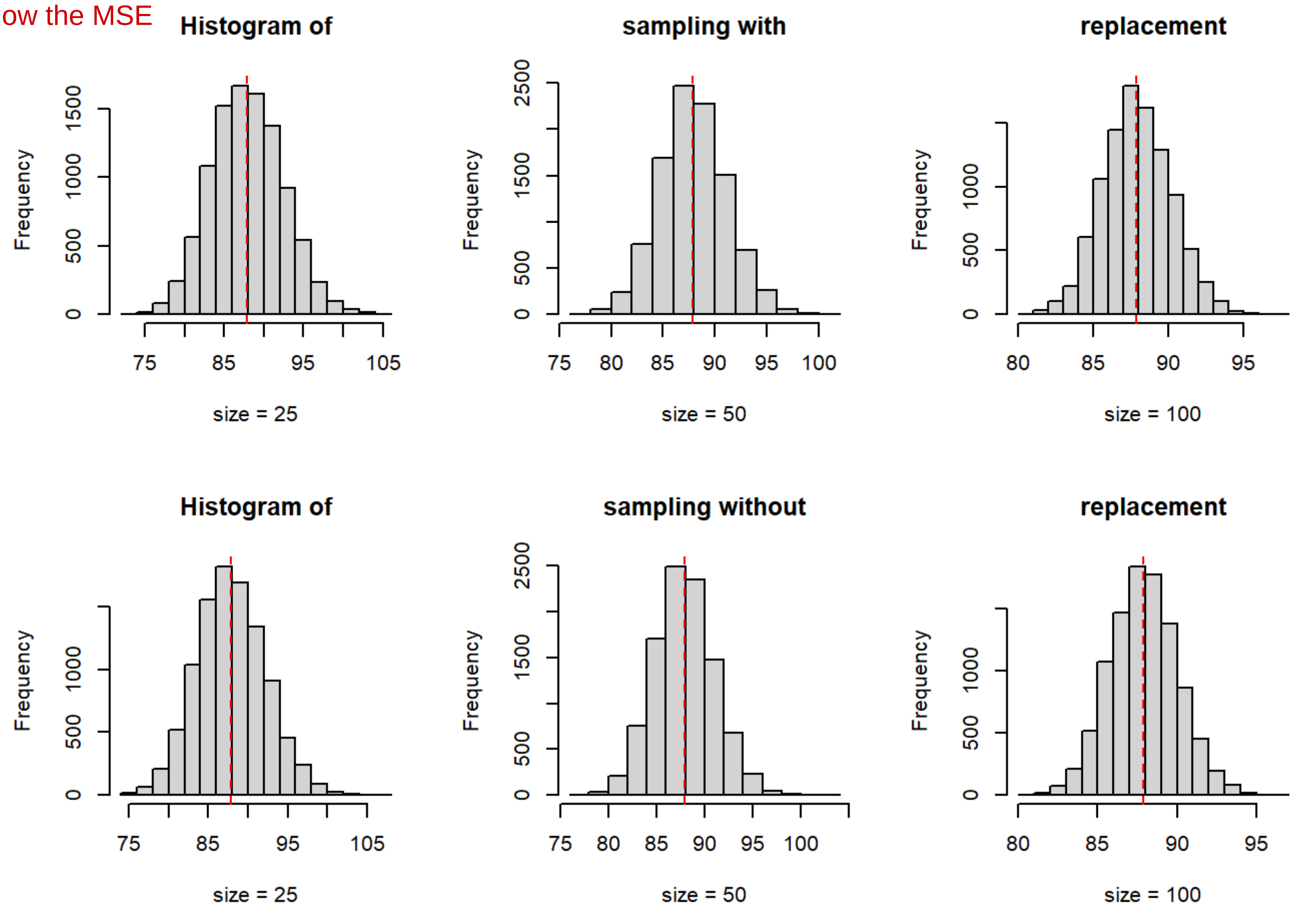
```r
sample_wr <- function(n)
{
  return(sample(population,n,replace = T))
}
#x <- sample_wr(n)
#mean(x)

sample_wor <- function(n)
{
  return(sample(population,n,replace = F))
}
#y <- sample_wor(n)
#mean(y)
```

(c) Repeat the experiment by taking n = 25, 50, 100. Based on your results, comment on the closeness of $\bar{Y}$ and μ under SRSWR and SRSWOR schemes, as n grows.

Given below are histograms of 10000 samples of mean of respective category(WR and WOR) and its given size. Here, red colored dotted line represent the mean value of the population which is equal to 87.90411

Show the MSE



As n grows, more values are clustering around the mean value of population or in other words variance decreases as n increases.

(d) Next consider the stratified sampling scheme with two strata: weekdays (stratum:1) and weekends (stratum:2). Considering $n_i \propto N_i$, find an estimate of μ.

Divide the population into two stratum, strata_1 (weekends) and strata_2 (weekdays)

```
dims <- c(length(strata_1),length(strata_2))

# function to sample from population taking inputs k the proportionality #constant , strata_1
and strata_2
strata_sample_1_wr <- function(k)
{
  value <- k*dims
  value <- floor(value) # for making dims integer

  sample <- c(sample(strata_1,value[1],replace = T),
               sample(strata_2,value[2]),replace = T)
  return(sample)
}
strata_sample_1_wor <- function(k)
{
  value <- k*dims
  value <- floor(value) # for making dims integer

  sample <- c(sample(strata_1,value[1],replace = F),
               sample(strata_2,value[2]),replace = F)
  return(sample)
}
```

Taking (for example) sample of 80 from strata_1 and accordingly from strata_2,

Mean Value for with replacement sampling

```
## [1] 88.17563
```

Mean Value for without replacement sampling

```
## [1] 87.5914
```

(e) Also, considering $n_i \propto N_i\sigma_i$, repeat the experiment, and compare the results.Here σ^2_i is the population variance of the i-th stratum.

```
# finding standard deviation of stratum, σ = (σ1,σ2)
sigma <- c(sd(strata_1), sd(strata_2))

# function to sample from population taking inputs k the proportionality #constant , strata_1
and strata_2

strata_sample_2_wr <- function(k)
{
  value <- k*sigma*dims
  value <- floor(value)

  sample <- c(sample(strata_1,value[1],replace = T),
               sample(strata_2,value[2], replace = T))
  return(sample)
}
strata_sample_2_wor <- function(k)
{
  value <- k*sigma*dims
  value <- floor(value)

  sample <- c(sample(strata_1,value[1],replace = F),
               sample(strata_2,value[2], replace = F))
  return(sample)
}
```

Taking (for example) sample of 80 from strata_1 and accordingly from strata_2,

Mean Value for with replacement sampling

```
## [1] 91.13736
```

Mean Value for without replacement sampling

```
## [1] 91.69076
```

(f) How will you compare the performance of the above four sampling schemes as n grows?

It can be compred using standard deviations.

The codes should be submitted in a separate file. Where are plots for stratified sampling?

```
# obv1 <- numeric(length = 290)
# obv2 <- numeric(length = 290)
# obv3 <- numeric(length = 290)
# obv4 <- numeric(length = 290)
#
# for(i in 1:290)
# {
#   sam1 <- numeric(length = 1000)
#   sam2 <- numeric(length = 1000)
#   sam3 <- numeric(length = 1000)
#   sam4 <- numeric(length = 1000)
#
#   for(j in 1:1000)
#   {
#     foo3 <- strata_sample_1_wr(i/sum(dims))
#     sam1[j] <- mean(foo3)
#     foo4 <- strata_sample_2_wr(i/sum(dims*sigma))
#     sam2[j] <- mean(foo4)
#     foo1 <- strata_sample_1_wor(i/sum(dims))
#     sam3[j] <- mean(foo1)
#     foo2 <- strata_sample_2_wor(i/sum(sigma*dims))
#     sam4[j] <- mean(foo2)
#   }
#   obv1[i] <- mean(sam1)
#   obv2[i] <- mean(sam2)
#   obv3[i] <- mean(sam3)
#   obv4[i] <- mean(sam4)
#
# }
#
# ind1 <- which(!is.na(obv1), arr.ind = T)
# ind2 <- which(!is.na(obv2), arr.ind = T)
# ind3 <- which(!is.na(obv3), arr.ind = T)
# ind4 <- which(!is.na(obv4), arr.ind = T)
#
# obv1 <- obv1[!is.na(obv1)]
# obv2 <- obv1[!is.na(obv2)]
# obv3 <- obv1[!is.na(obv3)]
# obv4 <- obv1[!is.na(obv4)]
#
# plot(x = ind1,y = obv1, type = "l" , col = "red",lwd = 2)
#
# lines(ind2, obv2, col = "blue", lty = "dashed")
# lines(ind3, obv3, col = "green", lwd = 2)
# lines(ind4, obv4, col = "purple", lwd = 2)
```